

Data Management for IDEAS



Illinois Dark Energy Astrophysical Survey



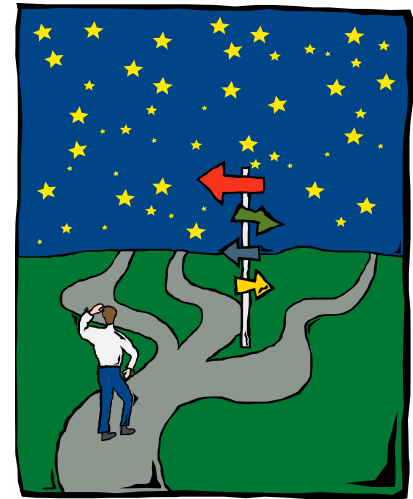
National Computational Science Alliance

National Center for Supercomputing Applications



Survey Data Management

- *This proposed survey is unlike previous surveys*
- **Large**
 - Order magnitude larger volume
 -
- **Point & Shoot**
 - Not Drift Scan
 - Typical “Thesis” Mode (Smith’s Talk)
 -
- **Science Drivers**
 - Primary Driver (Dark Energy)
 - Additional Drivers imply
 - Multiple Pipelines



Survey Data Management

- *This proposed survey is unlike previous surveys*

- No Existing Turnkey Solution

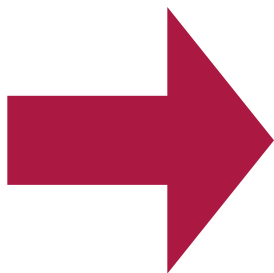


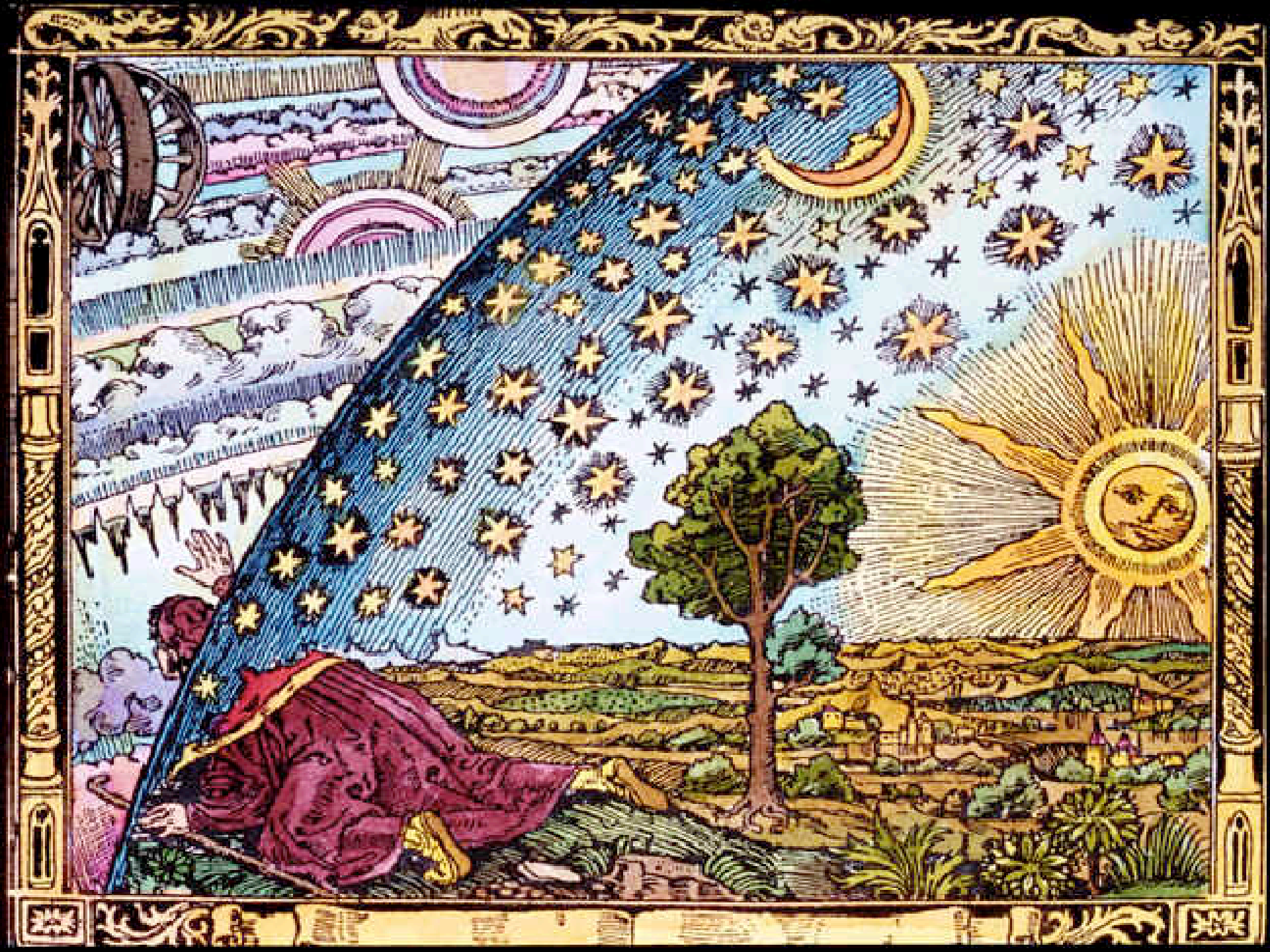
- But Existing tools & Algorithms

- SDSS Tools (Stoughton's Talk)
- Community tools

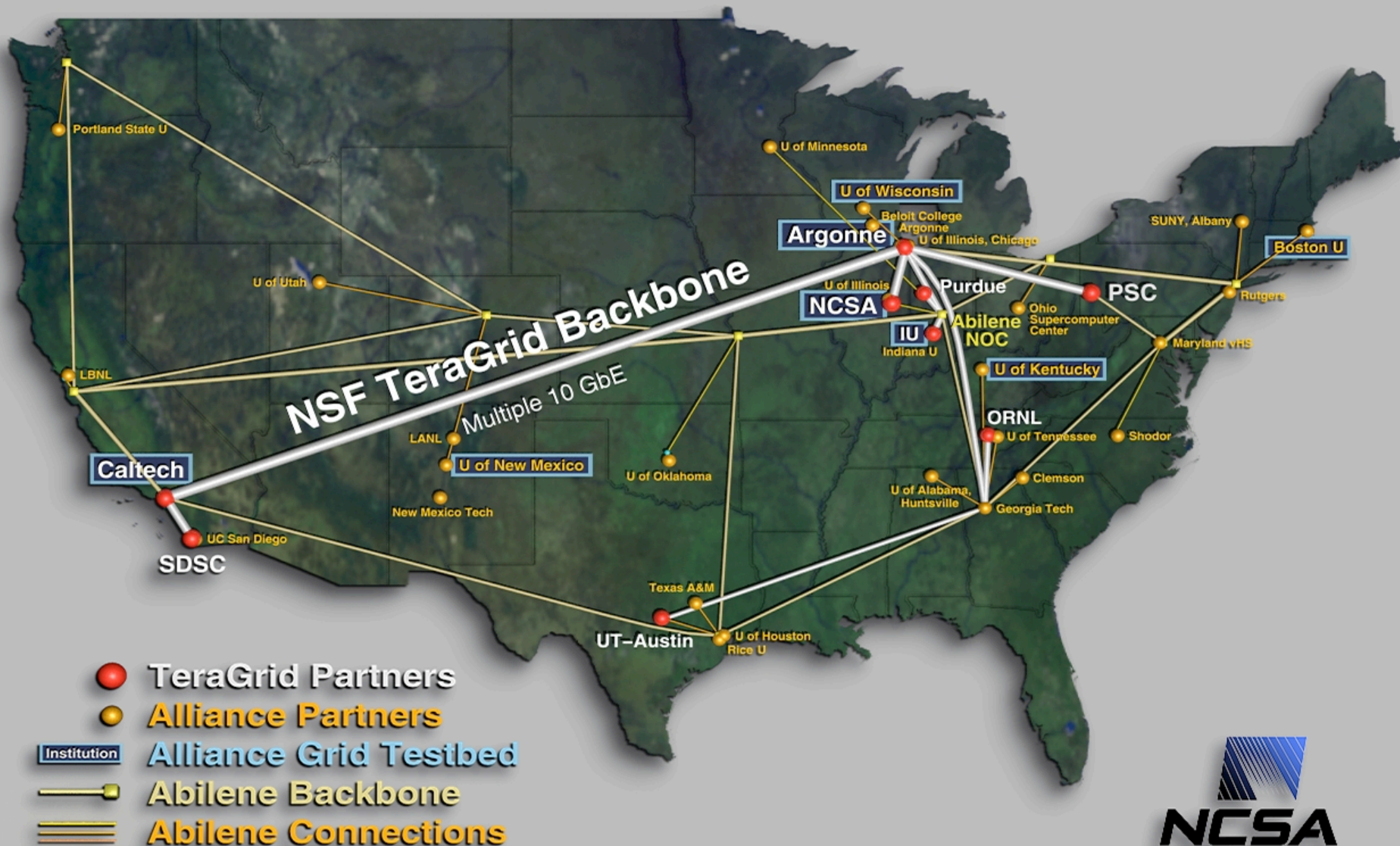


“Strawman” Overview





Alliance and ETF Sites



Extensible TeraGrid Facility (ETF)

Caltech: Data collection analysis

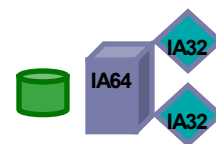
0.4 TF IA-64
IA32 Datawulf
80 TB Storage



LEGEND

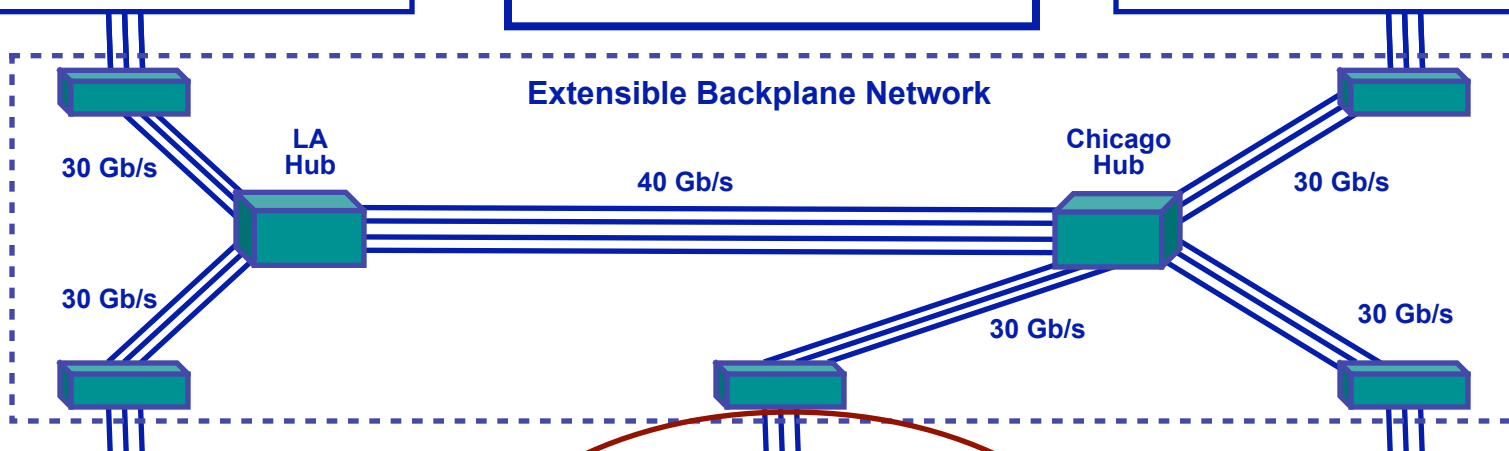


ANL: Visualization

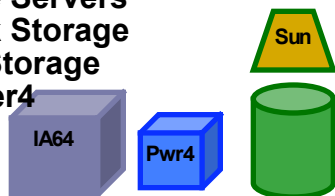


1.25 TF IA-64
96 Viz nodes
20 TB Storage

Extensible Backplane Network

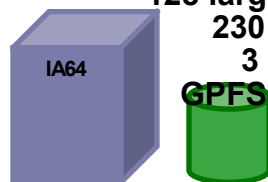


4 TF IA-64
DB2, Oracle Servers
500 TB Disk Storage
6 PB Tape Storage
1.1 TF Power4



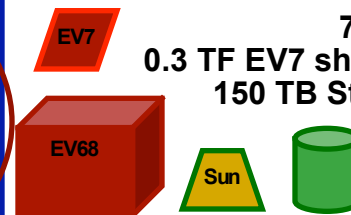
SDSC: Data Intensive

10 TF IA-64
128 large memory nodes
230 TB Disk Storage
3 PB Tape Storage
GPFS and data mining



NCSA: Compute Intensive

6 TF EV68
71 TB Storage
0.3 TF EV7 shared-memory
150 TB Storage Server



PSC: Compute Intensive

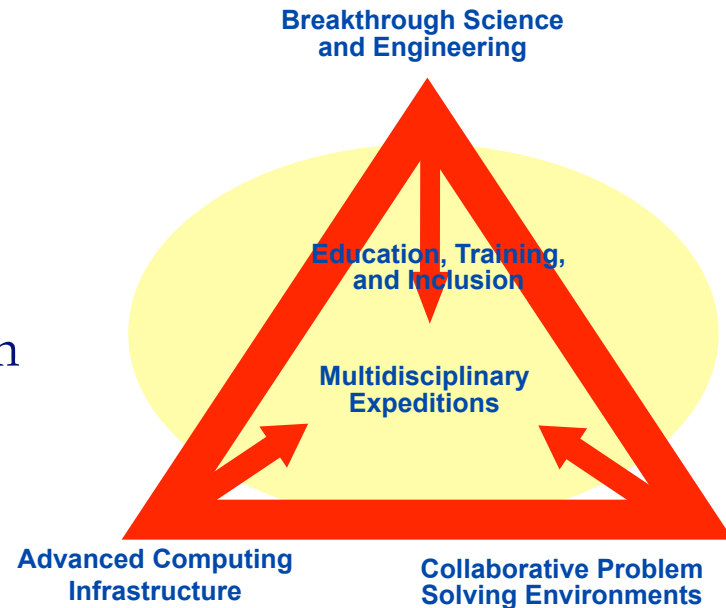


PACI Data Quest Expedition

PIs: Robert Brunner & Ian Foster

Leads: Miron Livny, Tom Defanti, Ray Plante,
Bob Wilhelmson, Jay Alameda

Achieve the technical advances,
technology deployment, and application
integration required to enable data-
intensive applications to harness
Alliance and TeraGrid resources



Astronomy major application area

Bulk Image Transfer

Bulk Image Processing

Bulk Data Mining

Data Distribution

CS Tools:

Globus

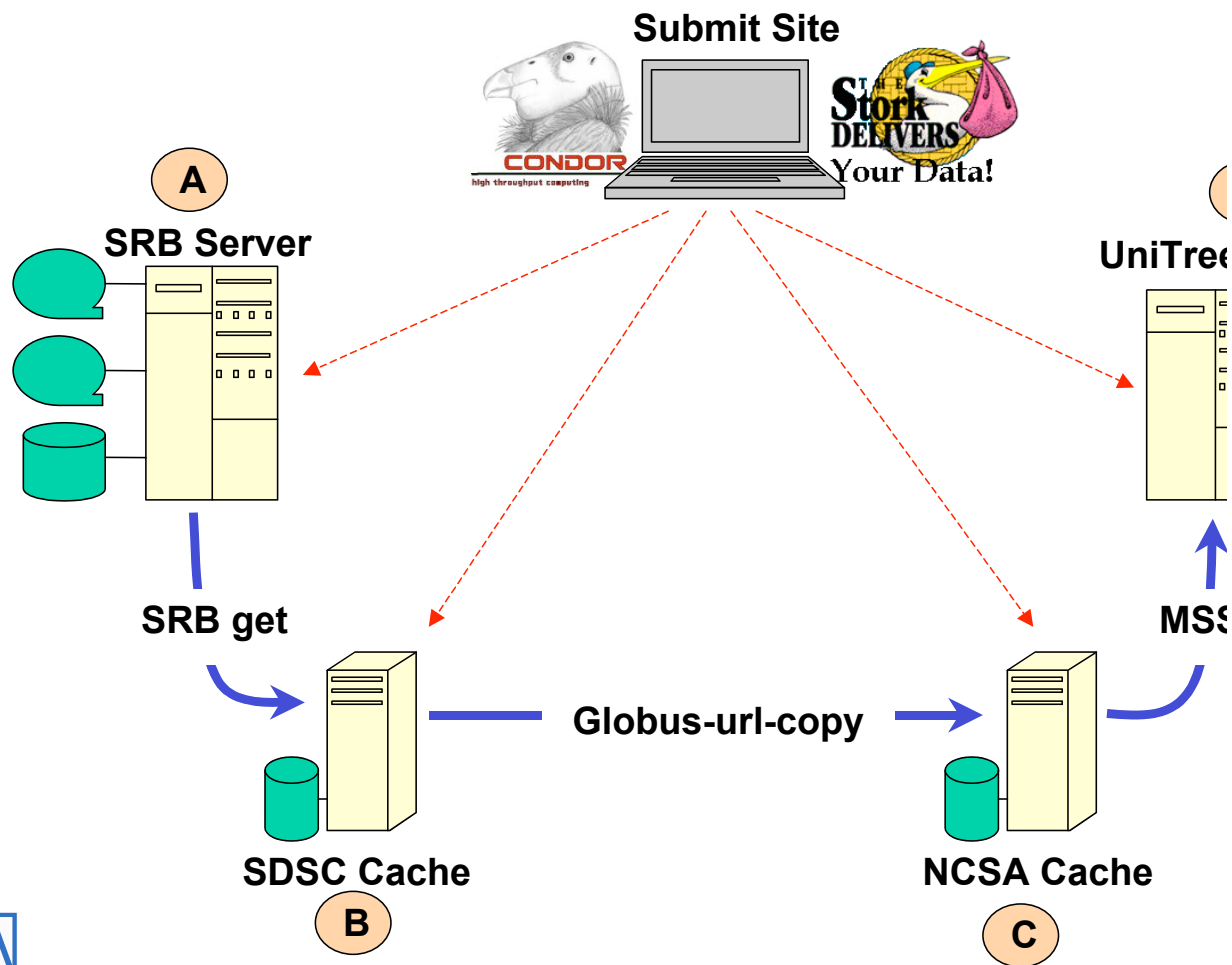
Chimera/VDT

Condor/Stork

MCS



Bulk Data Transfer



3 TB Data
Transferred

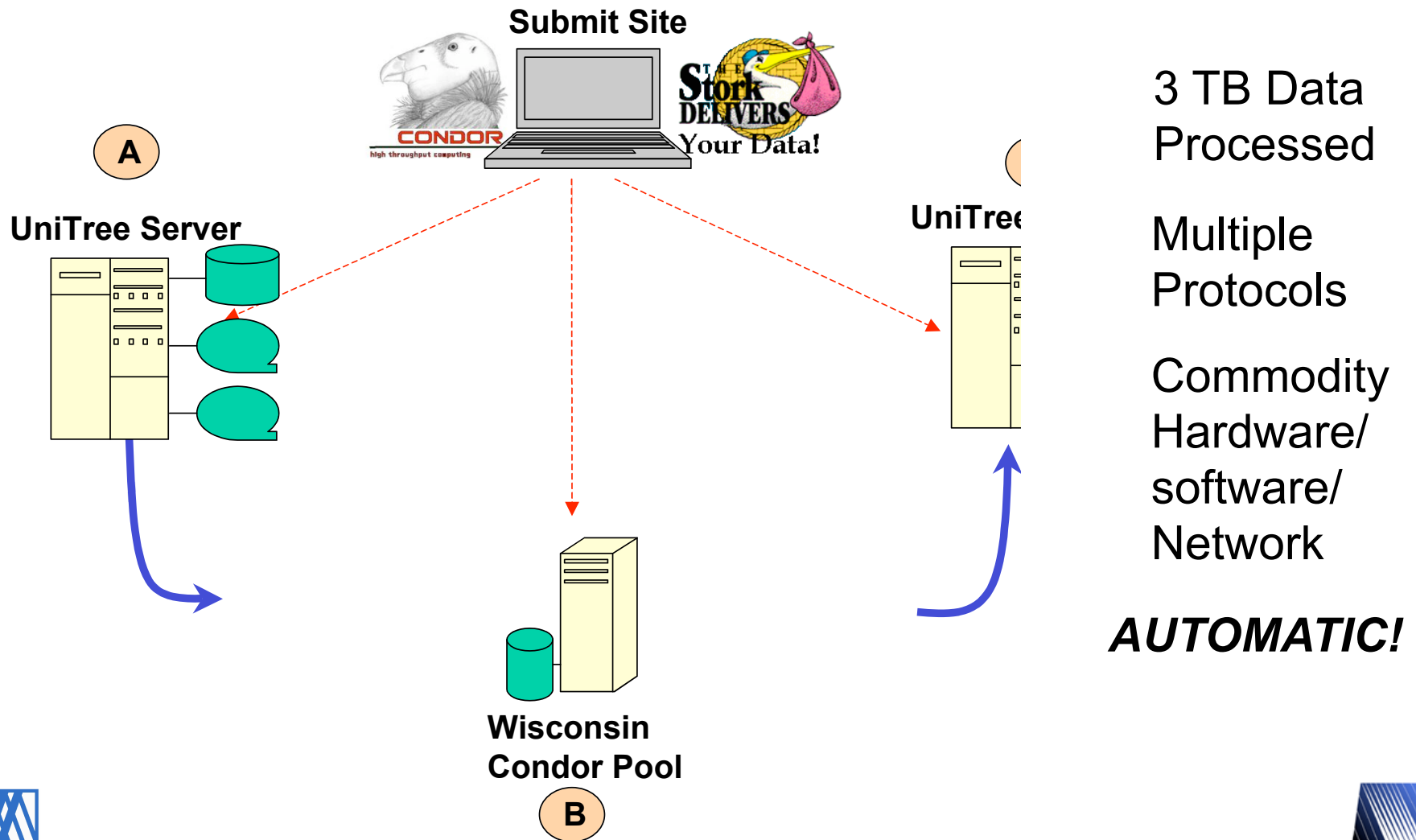
Multiple
Protocols

Commodity
Network

AUTOMATIC!

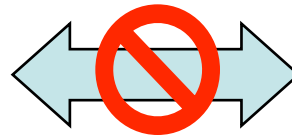
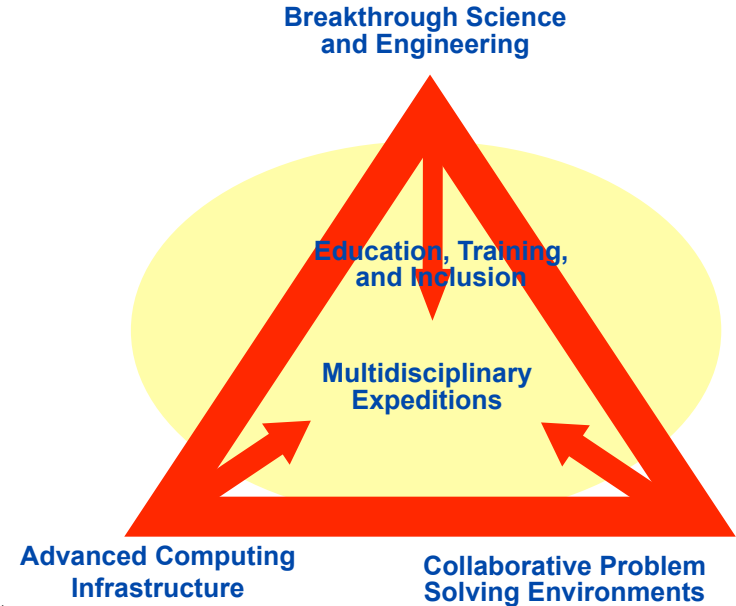


Bulk Processing



Data Management

- **Advocate for Alliance Model**
- **Single Lead Center**
 - *CPC Management Responsibility*
 - *Reduce support costs*
 - *CS Community Interface*
- **Multiple Partner Sites**
 - *Expand Intellectual Scope*
 - Pipeline/Simulation development
 - QA



NCSA Building



Growth & Stability!
Right now in 11+ buildings

Existing Hardware
Building
Hardened
Expansion(s)

- **Features**

- see www.ncsa.uiuc.edu/AboutUs/NCSABuildingWebcam.html
- State Funded
- \$30M, 142,000 square feet
- completion in 2005



Takeaways

- **Grid Tools Maturing**
 - Considerable experience & Partnerships
 -
- **Astrophysical Data Processing Grid Framework**
 - Wavelength neutral, Pipeline generic
 - Metadata management
 - In development now!
 -
- **Generic Data Management Experience**
 - Center-Level commitment
 - Multiple Astro/CS partners
 -
- **Staff Support (Both Hardware & Software)**
 - Processing, Networking, Archiving, Distribution

